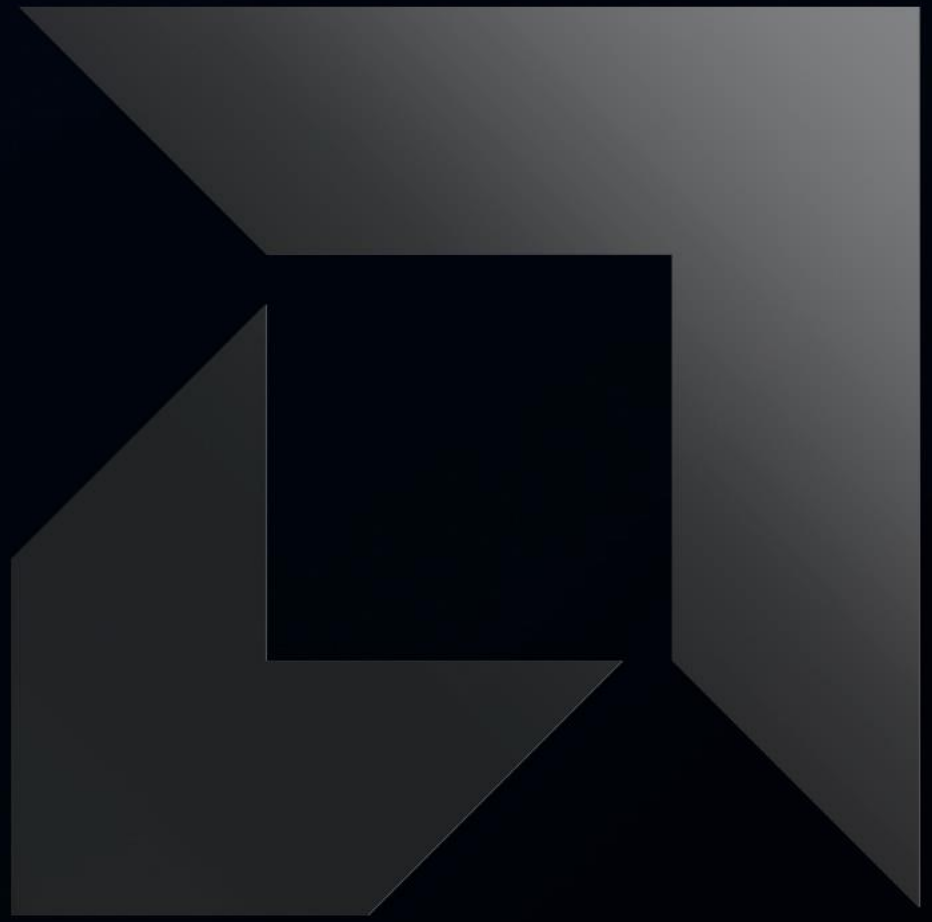


AMD GPUs in the UCX ecosystem

Edgar Gabriel

AMD
together we advance_

OUTLINE

- UCX
- UCC
- Open MPI
- MPICH
- Summary

AMD
together we advance_

UCX

ROCM SUPPORT IN UCX

- **ROCM/COPY**: data transfer between host and device memory within a single process
- **ROCM/IPC**: data transfer between device memories of different processes on the same node
- GPUDirect RDMA for communication between processes on different nodes
- Memory type detection of ROCm memory
- Memory hooks for ROCm memory allocations

AMD
together we advance_

RECENT WORK IN UCX

Revamping hardware agent selection in rocm/ipc component

- Limited number of devices might be visible to a process (`ROCR_VISIBLE_DEVICES`)

Changes to rocm memory detection

- Device vs. unified vs. host memory
- Simplified memory detection code at single location

Instinct MI250x support

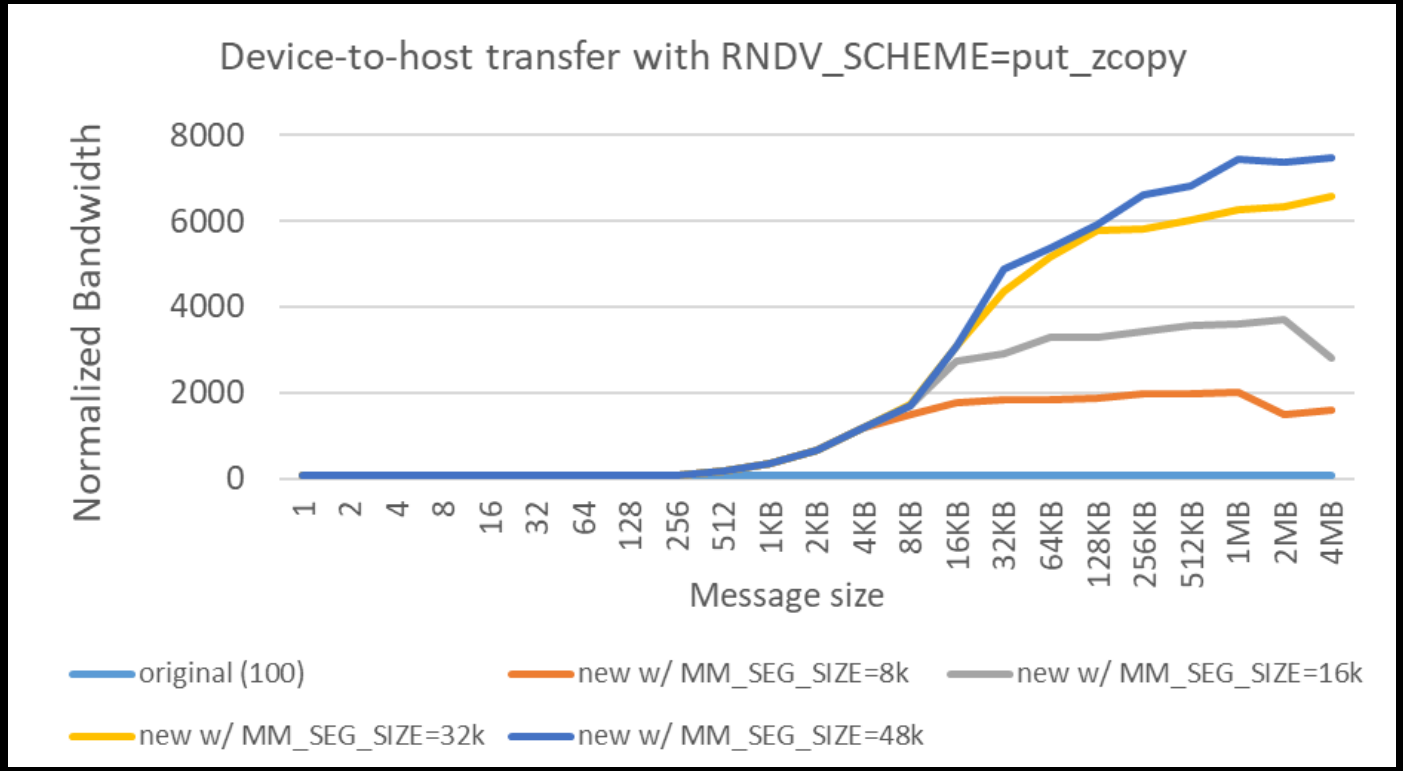
- Validation
- Tuning

AMD
together we advance_

RECENT WORK IN UCX (II)

New options for short operations in rocm/copy

- Host triggered copy operations (aka `memcpy`)
- Runtime library copy operations
- User controlled threshold to switch between both approaches



AMD
together we advance_

UCC

ROCM SUPPORT IN UCC

MC/ROCM:

- ROCm memory management

EC/ROCM

- Execute operations on ROCm memory
  - Data transfer
  - Reduction

TL/RCCL

- Execute collective operations using RCCL library

TL/UCP

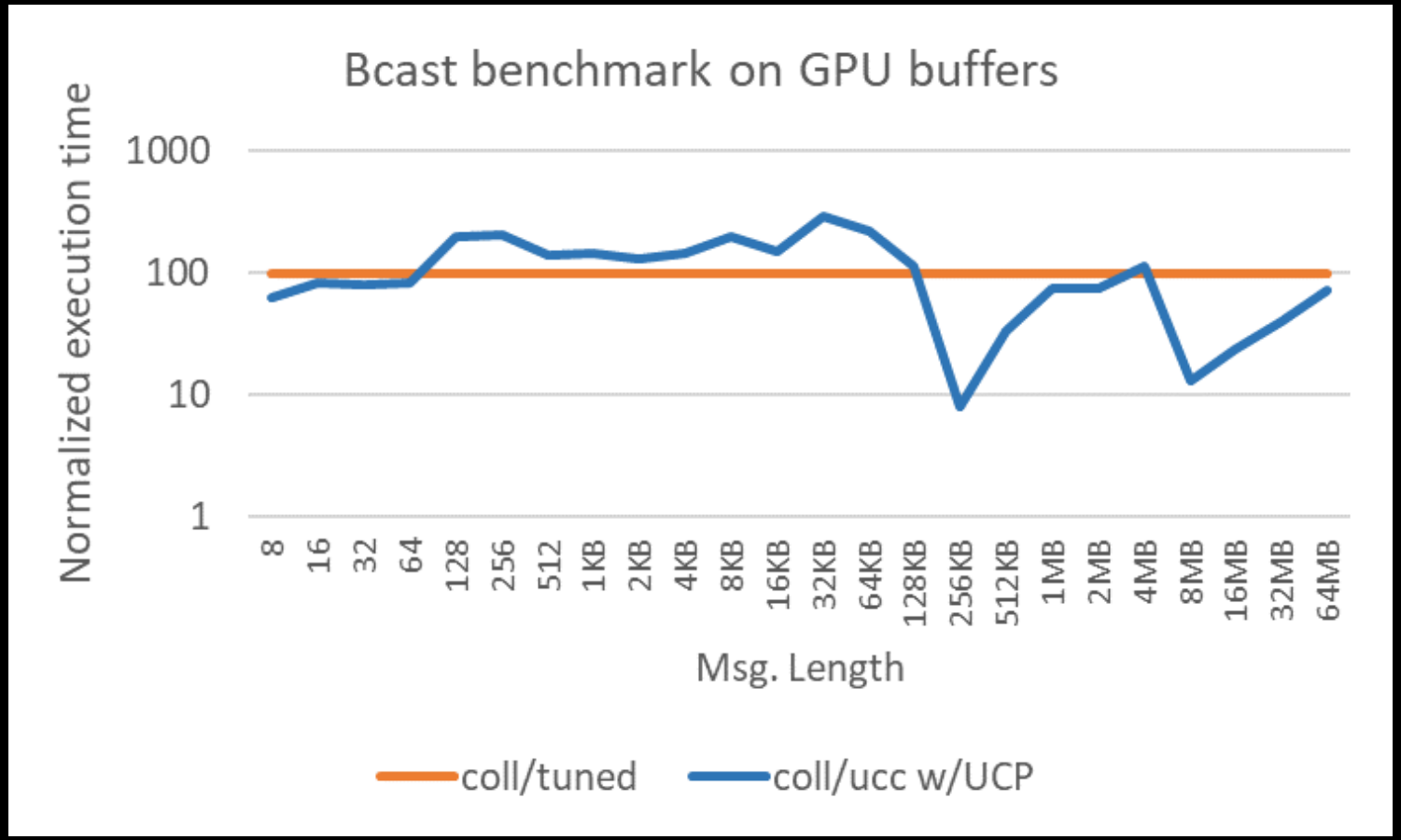
- Memory type agnostic
- Can invoke MC/ROCM and EC/ROCM operations internally

AMD
together we advance_

PRELIMINARY RESULTS: BCAST

- Open MPI main ~8/15/22
- UCX 1.13.0
- UCC pre 1.1

- 4 nodes
- 64 processes
- 64 MI250x GPUs
- 200Gb InfiniBand

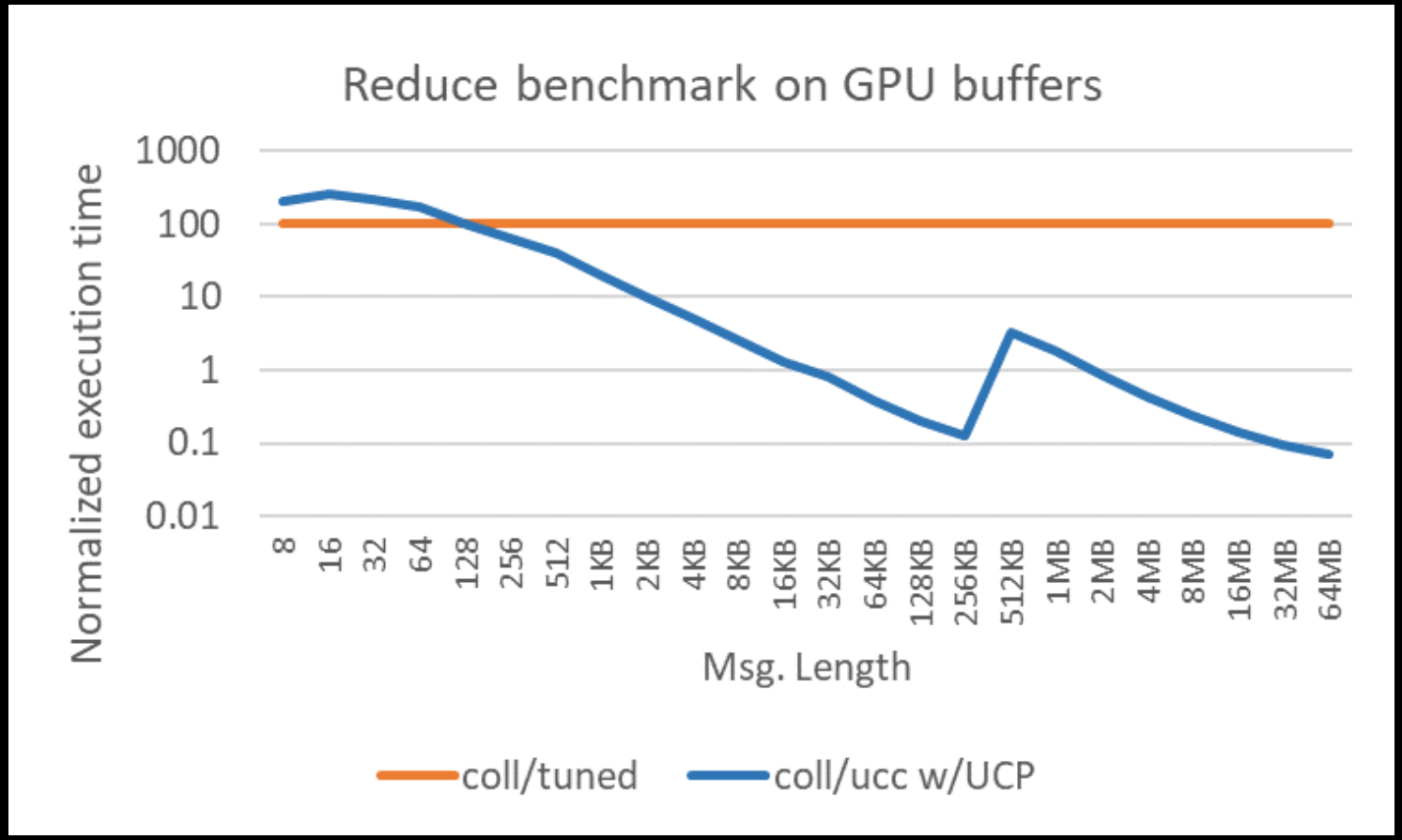


AMD
together we advance_

PRELIMINARY RESULTS: REDUCE

- Open MPI main ~8/15/22
- UCX 1.13.0
- UCC pre 1.1

- 4 nodes
- 64 processes
- 64 MI250x GPUs
- 200Gb InfiniBand



AMD
together we advance_

HOST BASED REDUCTION OPERATIONS ON GPU BUFFERS

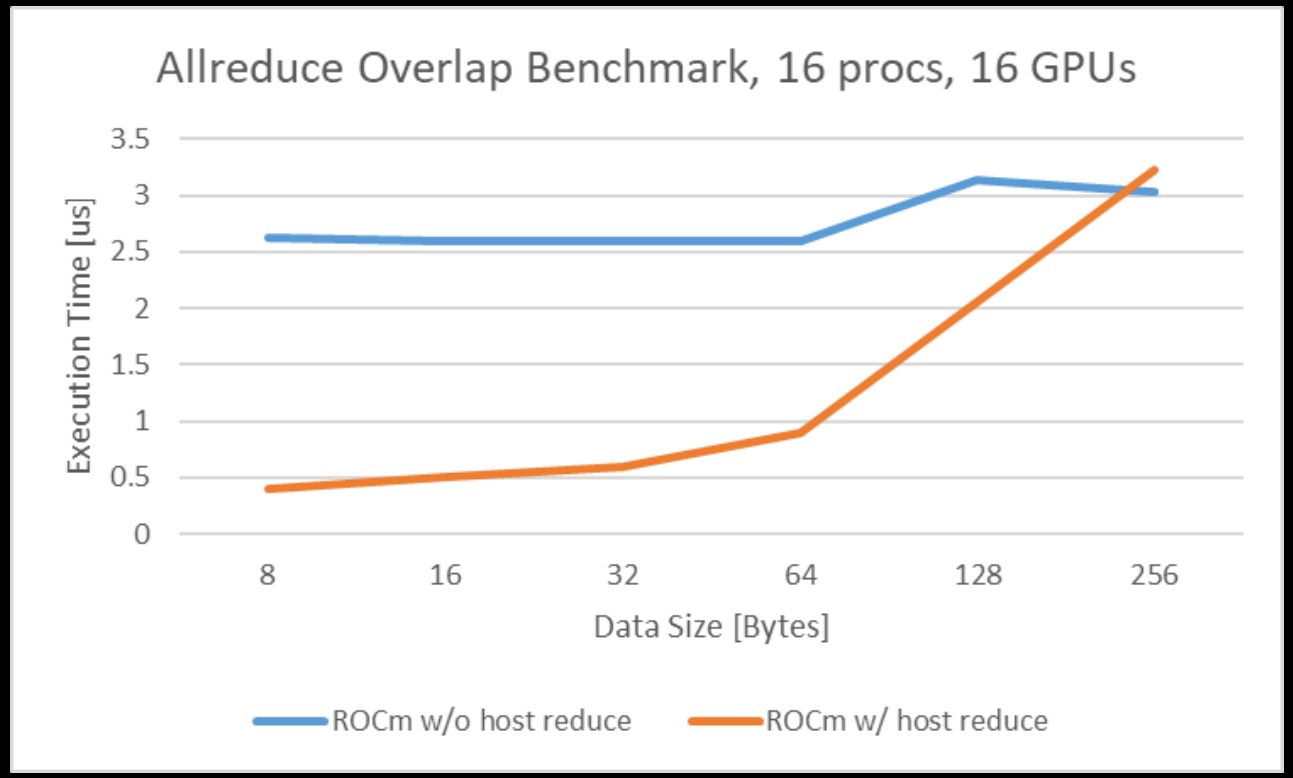
Extended EC/ROCM to perform reduction operations on host

- No device-to-host copy
- Avoids kernel launch(es)

➡ Lower latency

Use-cases

- Overlapping communication with computation
- Short buffers



AMD
together we advance_

Open MPI

OPEN MPI

- ROCm device memory supported through pml/ucx and osc/ucx
- Further ROCm support added in Open MPI in Spring 2022
  - Configure logic to detect ROCm software stack
  - ROCm memory type detection
  - Data transfer to/from ROCm device memory
    - Beneficial for non-contiguous derived datatypes

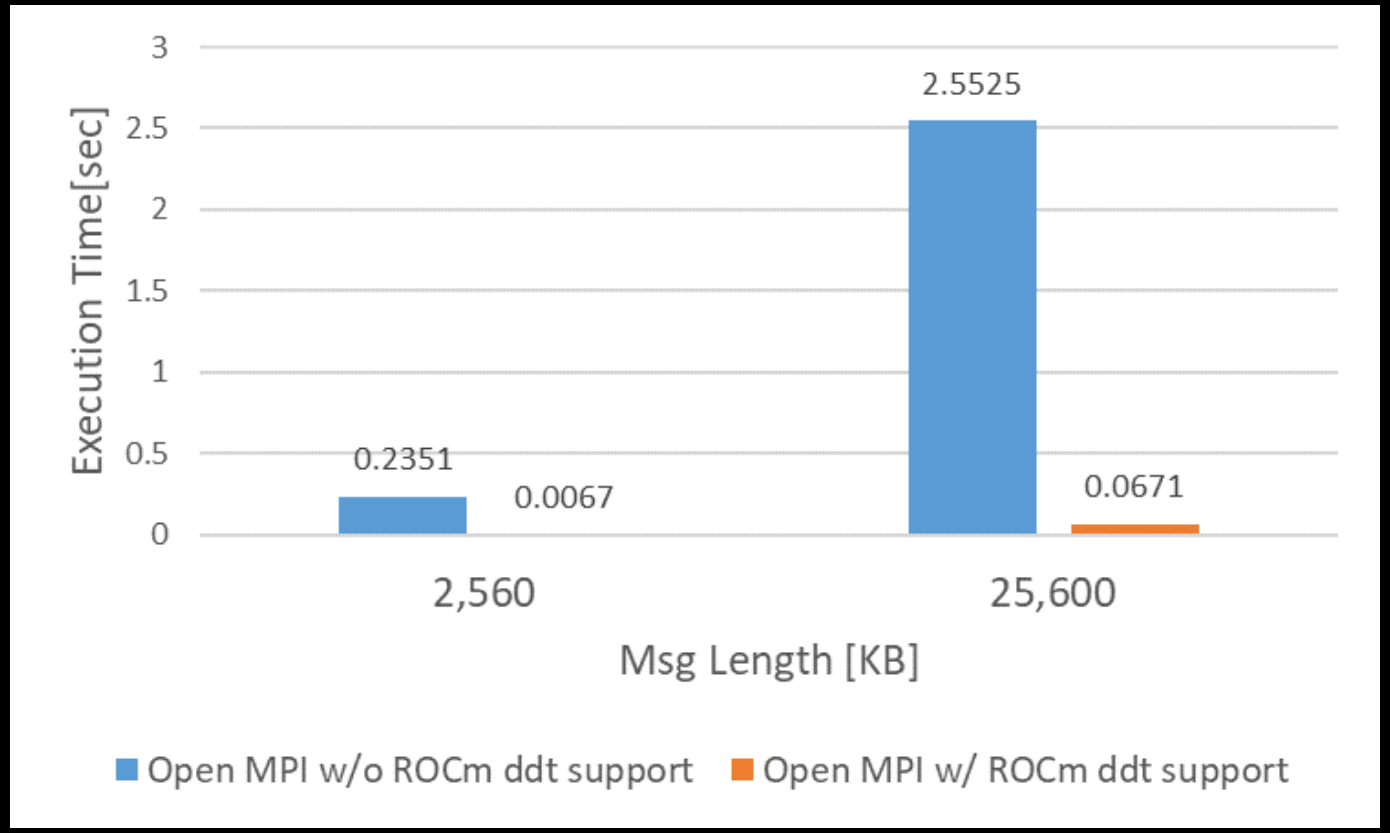
```
./configure --with-rocm=<ROCM_DIR>
            --with-ucx=<UCX_DIR>
            --with-ucc=<UCC_DIR>…
```

- Query availability of ROCm support in MPI library (`MPIX_QUERY_ROCM_SUPPORT`)
- HIP-MPI testsuite

AMD
together we advance_

OPEN MPI (II)

- Derived datatype benchmark using ROCm device memory with pml/ucx



AMD
together we advance_

OPEN MPI ACCELERATOR FRAMEWORK

- New framework introducing an abstraction layer for GPU utilization in Open MPI
  - Bulk of the work performed by Amazon
- Reduces GPU vendor specific code sections
- Open MPI can be compiled supporting multiple GPU vendors/APIs simultaneously
- GPU component selection at runtime
  - Only one GPU type per node allowed

AMD
together we advance_

OPEN MPI ACCELERATOR FRAMEWORK

- Components for CUDA and ROCm implemented
- Targeting Open MPI 5.0.0 release

AMD
together we advance_

MPICH

MPICH WITH UCX AND ROCM

ROCm devices supported in MPICH through UCX

```
./configure --with-ucx=<UCX_INSTALL_DIR>
            --with-hip=/opt/rocm
            --with-hip-sm=gfx90a
```

Difference in querying ROCm support

```
MPIX_QUERY_HIP_SUPPORT (MPICH)
MPIX_QUERY_ROCM_SUPPORT (Open MPI)
```

➡ Standardized interfaces required https://github.com/mpi-forum/mpi-issues/issues/580

AMD
together we advance_

SUMMARY

- UCX with ROCm support for point-to-point and RMA operations
- UCC with ROCm/RCCL support for collective operations
- Integration through Open MPI and MPICH for HPC applications

AMD
together we advance_

FUTURE WORK

- UCX
  - Support for new devices and architectures
  - Dealing with the large parameter space created by UCX
- UCC
  - Room for further improvements in performance of collective operations
  - Exploit device specific features
- Open MPI
  - Finalize accelerator framework support

AMD
together we advance_


AMD